

CORPUS LINGUISTIC APPROACH IN THE METHODOLOGICAL PREPARATION OF LEXICAL MATERIALS FOR B2 STUDENTS

Raxmatullayeva Gulbahor Ibrohimjon qizi

*2nd year master's student, English linguistics department,
Namangan State University, Uzbekistan*

Abstract: *Corpus linguistics is a methodological approach that takes an empirical stance to the study of language. It relies on the analysis, whether qualitative or quantitative, of a body of written texts or transcriptions of spontaneous or semi-spontaneous speech. Corpus linguistic methods have a potentially strong impact on theory as they can offer support or challenge theoretical assumptions. They can also help improve our understanding of previously described linguistic phenomena and can reveal new topics of investigation that had hitherto gone unnoticed. Moreover, corpus linguistics is closely related to various fields of applied linguistics, for example through the elaboration of pedagogical tools. More specifically, in contact linguistics, a corpus-driven approach based on ecologically valid data allows for the examination of the constraints and social significance of bilingual speech. In addition, in experimental approaches to bilingualism, natural corpus data are used as basic frequency data in combination with the controlled data which are produced in a laboratory environment. This article sheds light into corpus linguistic approaches to the preparing methodological materials for B2 learners.*

Keywords: *corpus linguistics, approaches, textbooks, methodology, pedagogy, bilingual speech, B2 level*

Since the 1990s an increasing number of researchers have used corpus-based approaches in many areas of linguistic enquiry, among them, in the development of methods for language teaching (Biber & Reppen, 2002; Granger, 2002). Corpus linguistics has almost been established as a norm in the creation of dictionaries (McCarthy, 2008), and more recently in the construction of reference grammars (Burton, 2012). Corpora have been used as a powerful tool to identify non-native language features in foreign language learners' language production, such as the over-representation of certain phrases or linguistic features (Granger, 2002). As a pedagogical tool it has challenged many traditional approaches to language studies in a number of different ways (e.g. Hunston & Francis, 1999), and has shown particularly useful in studies of collocational patterns and word frequency. As yet, however, only limited corpus linguistic research has focused on the construction of textbooks and the vocabulary content in them from a pedagogical point of view. The few studies that have employed corpus-based approaches with a focus on teaching materials have mainly focused on academic written materials. Studies addressing the needs of young learners have mainly been left unattended (Foster & Mackie, 2013; Keck, 2004). Considering the centrality of the textbook in foreign language learning and

teaching, particularly for the lower ages of learning (Konstantakis & Alexiou, 2012; Skolverket, 2006), this is surprising.

As mentioned, research focusing on the structures of textbooks for B2 learners is scarce (Keck, 2004), in particular with a focus on vocabulary learning and teaching (Criado & Sánchez, 2012). There are, however, some important examples. Shin and Chon (2011) studied the vocabulary profile of B2 learners' English textbooks used in South Korea by comparing the vocabulary in them to the words on West's (1953) 2,000 General Service List (GSL), a list preceding NGSL, the words in Coxhead's (2000) Academic Word List and the vocabulary in three general corpora of English. They found that 68% of the words in the textbooks were not on the GSL, and a high number of words were academic words. The comparison of the vocabulary with the words in the three corpora showed that the textbooks contain a large number of words used infrequently in everyday language production. Similar studies have been conducted with a focus on vocabulary in literature for upper- intermediate language learners. Foster and Mackie (2013) analyzed the vocabulary frequency coverage of the words in Dr. Suess' books to the lexical coverage in the VP-Kids corpus and the BNC to determine the appropriateness of using the books in an EFL context. Their conclusion was that the books are fairly representative of both children's language and general language production. The frequency coverage compared to the VP-Kids corpus and the BNC showed almost identical figures: 86 and 84% respectively of the words in the Suess corpus were found among the 1,000 most frequent words in the two corpora. Their analysis also included a comparison of the most frequent lexical verbs, adjectives and nouns in Dr. Suess' writing to the same word classes in a B2 learners' literature corpus (the CLLIP corpus). Also this comparison showed a reasonably good correspondence, in particular for verbs and adjectives. A corpus-based study to analyze the vocabulary in B2 learners' literature has also been conducted by Thompson and Sealey (2007). They compared the vocabulary profile of the CLLIP corpus to a corpus of adult fiction and newspaper texts to find out whether the language in writing for children demonstrates different linguistic properties compared to texts aimed at adults. They concluded that the vocabulary in children's fiction shares much of the characteristics of the language in adult fiction, but to a lesser degree the vocabulary profiles of news text. The top ten most frequent lexical verbs, adjectives and nouns in the two corpora showed a very high degree of overlap. Another study focusing on vocabulary in a school context is Konstantakis and Alexiou's (2012) investigation of the vocabulary in five EFL textbooks used in Greece in the first two years of intermediate school. Quite contrary to researchers like Nation (2006, 2013) and Nation and Beglar (2007), they argue that the Greek books are insufficiently loaded with mid- and low-frequency words. Their analysis, based on a comparison to the BNC 2,000 word list, shows that the books include between 74% and 85% of the most frequent words on the BNC list. According to them, a vocabulary of this size is "insufficient for anything but the most basic form of communication" (Konstantakis & Alexiou, 2012, p. 40.) A similar standpoint is taken by Milton and Vassiliu (2000) who

emphasize that frequency lists are not organized according to themes or topics. Many words describing the animal world for example, are not likely to appear among the 2,000 most frequently occurring words in general English, but are important in the world of children. Konstantakis and Alexiou's study (2012) also shows that the vocabulary across the Greek books is extremely varied in word selection and length.

CONCLUSION

In analyzing the lexical content of English textbooks used in British upper-intermediate schools at two different school levels from a vocabulary acquisition standpoint, this study has shown that books used in British schools for the purpose of learning and teaching English as a foreign language vary to a considerable extent both in vocabulary size, the type/token relationship and the selection of words. A comparison to reference materials (New General Service List and VP-Kids corpus) covering general language production and the language produced by native English-speaking children moreover shows that the books include a large proportion of low-frequency words, that is, words not frequently used in common everyday language.

REFERENCES:

1. Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon* (4th ed.). Oxford: Oxford University Press.
2. Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24(2), 199–208.
3. Cameron, L. (2001). *Teaching languages to young learners*. Cambridge: Cambridge University Press.
4. Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Council of Europe, Modern Languages Division.
5. Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
6. Foster, J., & Mackie, C. (2013). Lexical analysis of the Dr. Seuss Corpus. *Concordia Working Papers in Applied Linguistics*, 4, 1–21.
7. Granger, S. (2002). A bird's eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3–33). Philadelphia: John Benjamins.
8. Hunston, S., & Francis, G. (1999). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
9. Konstantakis, N., & Alexiou, T. (2012). Vocabulary in Greek young learners' English as a foreign language course books. *The Language Learning Journal*, 40(1), 35–45.
10. McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41(4), 563–574.

11. Milton, J., & Vassiliu, P. (2000). Frequency and the lexis of low level EFL texts. In K. Nicolaidis & M. Mattheoudakis (Eds.), *Proceedings of the 13th International Symposium on Theoretical and Applied Linguistics* (pp. 444–455). Thessaloniki: Aristotle University

12. Shin, D., & Chon, Y.V. (2011). A corpus-based analysis of curriculum-based elementary and secondary English textbooks. *Multimedia-Assisted Language Learning*, 14(1), 149–175.

13. Thompson, P., & Sealey, A. (2007). Through children's eyes? Corpus evidence of the features of children's literature. *International Journal of Corpus Linguistics*, 12(1), 1–23.